

基于 DLT-Kmedoids 算法的用电负荷聚类分析

陈苏豫¹, 顾亦然^{1,2}, 张腾飞¹

(1. 南京邮电大学 自动化学院、人工智能学院, 江苏 南京 210023;

2. 南京邮电大学 智慧校园研究中心, 江苏 南京 210023)

摘要:针对高校用电负荷中传统聚类算法直接应用于时间序列聚类效果准确性较低的问题,提出一种融合 DTW 距离、LB_Keogh 距离以及时间窗口的 DLT-Kmedoids 算法,以提高聚类算法应用于时间序列的准确性以及算法效率。DLT-Kmedoids 算法使用 DTW 计算时序数据之间的距离取代传统的欧氏距离度量方式,提高了相似性度量算法精度,同时也提高了聚类算法的准确性和复杂度,引入 LB_Keogh 距离在计算 DTW 距离之前过滤掉大部分不可能是最优匹配序列的序列,对于剩下的序列再使用 DTW 逐个比较,进一步降低算法的复杂度。最后结合高校建筑用电负荷时间序列数据进行分析,通过与主流聚类算法进行比较,表明该算法对于高校用电负荷数据的聚类任务,能够更准确地识别相似的负荷模式,并以更高的效率进行聚类分析。

关键词:用电负荷数据;动态时间弯曲;LB_Keogh;聚类;用电模式

中图分类号:TP181;TM743

文献标识码:A

文章编号:1673-629X(2024)04-0205-07

doi:10.20165/j.cnki.ISSN1673-629X.2024.0031

Power Load Clustering Analysis Based on DLT-Kmedoids

CHEN Su-yu¹, GU Yi-ran^{1,2}, ZHANG Teng-fei¹

(1. School of Automation and School of Artificial Intelligence, Nanjing University of Posts and

Telecommunications, Nanjing 210023, China;

2. Smart Campus Research Center, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: Aiming at the problem of low accuracy in directly applying traditional clustering algorithms to time series clustering in the electricity load of universities, a DLT-Kmedoids algorithm combining DTW distance, LB_Keogh distance and time window is proposed to improve the accuracy and efficiency of clustering algorithm applied to time series. The DLT-Kmedoids algorithm uses DTW to calculate the distance between time series data instead of traditional Euclidean distance measurement, improving the accuracy of similarity measurement algorithms and also improving the accuracy and complexity of clustering algorithms. LB_Keogh distance is introduced to filter out most sequences that are unlikely to be optimal matching sequences before calculating DTW distance, and DTW is used to compare the remaining sequences one by one to further reduce the complexity of the algorithm. Finally, we analyze the time series data of electricity consumption in university buildings, and compare it with mainstream clustering algorithms. It is showed that the proposed algorithm can more accurately identify similar load patterns and perform clustering analysis with higher efficiency for the clustering task of electricity consumption data in universities.

Key words: power load data; dynamic time warping; LB_Keogh; clustering; power consumption mode

0 引言

近年来,随着智能电网的发展,大量智能化采集终端的不断接入、信息化系统的不断推广和智能楼宇等工程的不断推进,极大丰富了用电管理的数据基础。电力系统运行中的数据来源于各个环节,智能终端^[1]采集到的数据内容更加丰富,在高校用电管理系统信

息化改造后,每日可以采集各建筑物用电数据达百万条,并且数量还在逐年增长。但是,海量的信息导致了电力数据的冗余与杂乱。因此,如何从用电数据中提取有效信息,分析用户用电模式,为用电决策提供参考,已被认为是电网领域的重要研究内容^[2]。电力负荷曲线聚类技术是分析用户用电特征的重要手段,也

收稿日期:2023-06-02

修回日期:2023-10-09

基金项目:国家自然科学基金(62073173)

作者简介:陈苏豫(1999-),男,硕士研究生,研究方向为负荷建模;通信作者:顾亦然(1972-),女,教授,博士,CCF会员(E200012844M),研究方向为复杂网络、大数据处理等。

是电力数据挖掘技术的基础。管理者通过挖掘出的典型负荷曲线,可以有针对性地制定分时电价,对建筑物用电曲线进行去峰填谷,对用电用户精细分类和用电管理有着重要的意义。

时间序列聚类分析算法^[3]被广泛应用于用户用电行为特征提取。文献[4]对去除了异常负荷的时间序列数据直接使用 Kmeans 聚类算法来提高负荷预测结果的准确性。文献[5]首先对时间序列数据根据特征标签分开,然后归一化处理实现降维,再基于谱聚类对风电场内风机出力特性曲线进行聚类,并且建立功率预测模型。层次聚类先计算样本之间的距离。每次将距离最近的点合并到同一个类,然后计算类与类之间的距离,将距离最近的类合并为一个大类。不停的合并,直到合成了一个类。文献[6]使用用户数据的协方差表征用户用电特征,然后使用层次聚类对用户特征进行聚类,最终实现对用户用电负荷的预测。基于密度的算法是基于相对范围内的密度进行聚类,可以有效过滤噪声数据,以设置的密度参数作为终止条件。文献[7]引入用户画像表征用电负荷时间序列实现降维,然后使用密度聚类算法结合熵权法构建指标间的关系,最终实现对用电时间序列的聚类。除此以外,还有通过选取根据时间序列的数据特征定制合适的距离度量函数,实现原始时间序列数据聚类^[8]。文献[9]使用欧氏形态距离代替欧氏距离有效区分了负荷曲线的变化特征。文献[10]使用粒计算定义粒度距离并结合欧氏距离度量曲线相似性,提升了聚类的质量。文献[11]使用 DTW-Kmedoids 算法对电压曲线实现聚类分析。这种方法大大提升了时间序列聚类的准确率,但是也大大提升了算法复杂度和计算时间。最近还有学者提出将深度学习融入时间序列聚类中。文献[12]使用深度卷积嵌入聚类算法提取典型负荷曲线,然后用一维卷积自编码器提取特征,相较于传统算法,聚类效果得到了提升。文献[13]在聚类算法中引入近邻图优化计算的时间与空间复杂度,实现用户用电特征的准确快速提取。

综上所述,目前大部分对于用电负荷曲线的聚类方法主要分为直接聚类和间接聚类。直接聚类是指直接通过修改度量距离对用电负荷曲线进行聚类,这种方法应用于高维时间序列时,算法复杂度高、计算时间长。间接聚类根据用电负荷曲线构建用户用电特征,然后用传统聚类算法对特征指标进行聚类,虽然大大降低了计算量,但是忽略了时间序列含有的大量信息。融合了深度学习算法的智能聚类算法虽然能够快速提取特征,但是过度依赖算力,并且在聚类准确性上要差于传统算法。直接聚类算法复杂度高,间接聚类忽略了曲线包含的时间信息。综合考虑时间信息与算法复

杂度,该文提出结合 DTW^[14]距离、LB_Keogh 距离^[15]与时间窗口的 DLT-Kmedoids 算法,降低了时间序列曲线误匹配率以及算法复杂度。

1 DLT-Kmedoids 算法

该文提出的 DLT-Kmedoids 算法首先进行数据预处理,使用均值插补法填补数据空缺值,使用均值平滑法减少数据噪声。然后使用 DTW 距离代替欧氏距离计算时序数据之间的距离,提高相似性度量的准确性。通过引入 LB_Keogh 距离将非最优匹配序列过滤从而降低误匹配率以及计算时间,引入滑动时间窗口将算法计算窗口规定在某一时间段内进一步降低运算时间。最终通过多次迭代寻找评价指标最优将样本归为 K 类。

算法流程如图 1 所示。

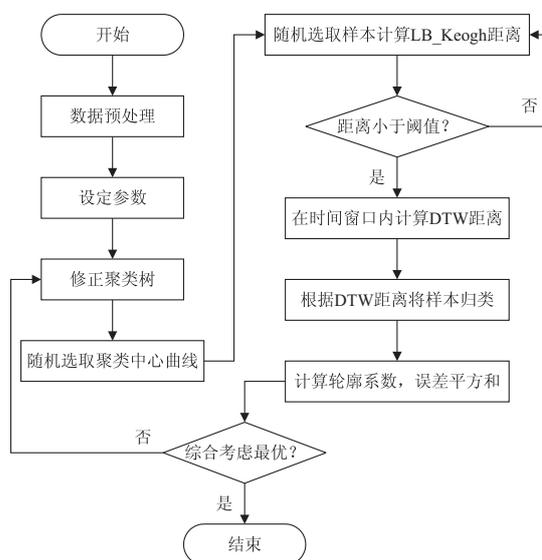


图 1 算法流程

1.1 数据预处理

实验对象选择的是南京某高校电力数据采集系统所记录的用电负荷数据。在电力数据采集系统记录的海量数据中存在着大量不完整、不一致的数据,这些数据会影响算法执行的准确率,导致进行数据分析时会产生误判,因此通过数据预处理得到较为完整的数据集尤为重要。

1.1.1 插值

电力数据采集系统每隔 15 分钟对电力用户进行检测,一天共有 96 个采样点。电力数据采集系统在采集电量数据的过程中由于数据传输、装置故障、线路等问题会产生缺失值,会对算法执行的结果产生较大影响。对此,该文采取均值插补法^[16]补全缺失值。假设一天完整负荷数据集为 $T = \{t_1, t_2, \dots, t_{96}\}$, $T_{\text{miss}} = \{t_m, t_{m+1}, \dots, t_{m+n}\}$ 为一段连续缺失值集合。缺失值按公式 1 和公式 2 进行插补。

$$t_{m+n} = \begin{cases} 0, m = \{1, 96\}, n = 0 \\ t_{m-1} + t_{up}, m + n < 96, n \neq 0 \end{cases} \quad (1)$$

$$t_{insert} = (n + 1) * \frac{t_{m+n+1} - t_{m-1}}{n + 2} \quad (2)$$

式中, m 为第一个缺失值在完整数据集 T 中的位数, n 为索引值, t_{insert} 为均值插补计算值。

1.1.2 数据平滑处理

电力数据采集系统每隔 15 分采集的数据点粒度过小,数据维度高导致算法复杂度过高,因此,该文将每小时内的 4 个采样点数据进行累加。最终合并为一天 24 个采样点。经过数据合并、插值后的一天完整负荷数据集为 $T_{24} = \{t_1, t_2, \dots, t_{24}\}$ 。电力数据采集系统在实际运行时由于受到通信中断、软硬件故障、信号干扰等影响会使得负荷数据产生噪声^[17],导致负荷曲线出现较大波动,从而影响聚类结果。因此,该文采用均值平滑法对用电负荷曲线进行平滑处理,使得负荷曲线更为平滑,从而进一步突出曲线形状,反映出曲线的总体趋势。

数据平滑算法如下:

Step1:输入包含 24 个采样点的数据集合,平滑次数 o ,平滑阈值 tw 。

Step2:计算连续两个采样点差值的绝对值,记为 dv 。

Step3:判断插值 dv 是否小于平滑阈值 tw 。如果小于则替换为两者平均值,否则采样点数值保持不变。

Step4:根据平滑次数 o ,循环执行 Step3。

Step5:输出数据平滑后的曲线集合 T_{24} 。

1.1.3 数据归一化

由于每栋建筑物的用电负荷量级存在差异,为了避免小数据淹没于大数据,采用 min-max 归一化按公式 3 将不同建筑物的用电量数据归在 $[0, 1]$ 之间。

$$E_{bi} = \frac{e_{bi} - e_{min}}{e_{max} - e_{min}} \quad (3)$$

式中, E_{bi} 为归一化后的第 bi 栋楼用电量数据, e_{bi} 为归一化前的第 bi 栋楼的用电量数据, e_{min} 和 e_{max} 分别为第

i 栋楼用电量的最小值和最大值。

1.2 使用 DTW 距离代替欧氏距离

在语音识别领域^[18],有关学者提出动态弯曲距离 (Dynamic Time Warping, DTW) 作为一种新的相似性度量方法,通过调节时间序列采样点的对应关系,寻找时间序列的最佳匹配路径,可以更加有效地度量时间序列间的距离。

假设有两个时间序列 $A = \{a_1, a_2, \dots, a_y\}$, $B = \{b_1, b_2, \dots, b_z\}$, 当 $y = z$ 时可以直接用公式 4 进行欧氏距离度量,当 $y \neq z$ 时,则需要先将 A 和 B 对齐。

$$d(A, B) = \sqrt{(a_1 - b_1)^2 + \dots + (a_y - b_z)^2} \quad (4)$$

式中, $d(\cdot)$ 为样本间的欧氏距离。欧氏距离度量图如图 2(a) 所示。

将时间序列 A 和序列 B 中的元素组成一个 $y \times z$ 的矩阵,序列 A 和 B 中对应两点之间的距离组成穿越路径 $W = \{w_1, w_2, \dots, w_{pi}\}$,穿越路径中的第 pi 个点记作 $w_{pi}(a_y, b_z)$ 。其中 w_{pi} 为第 pi 个穿越路径, a_y 为序列 A 中第 y 个值, b_z 为序列 B 中第 z 个值。

穿越路径有如下约束:

$$(1) w_1 = \{a_1, b_1\}, w_{pi} = \{a_y, b_z\};$$

$$(2) \text{如果 } w_{pi} = \{a_y, b_z\}, w_{pi+1} = \{a'_y, b'_z\}, \text{那么 } a'_y - a_y \leq 1;$$

$$(3) \text{如果 } w_{pi} = \{a_y, b_z\}, w_{pi+1} = \{a'_y, b'_z\}, \text{那么 } a'_y - a_y \geq 0, b'_z - b_z \geq 0。$$

基于以上约束,根据动态时间规整,可以得到 DTW 距离迭代公式,如公式 5 和公式 6 所示。

$$D(A, B) = \sqrt{\sum w_{pi}(a_y, b_z)} \quad (5)$$

$$w_{pi}(a_y, b_z) = d(a_y, b_z) + \min \begin{cases} D(a_{y-1}, b_{z-1}) \\ D(a_{y-1}, b_z) \\ D(a_y, b_{z-1}) \end{cases} \quad (6)$$

式中, $D(\cdot)$ 为样本间的 DTW 距离。DTW 距离度量如图 2(b) 所示。

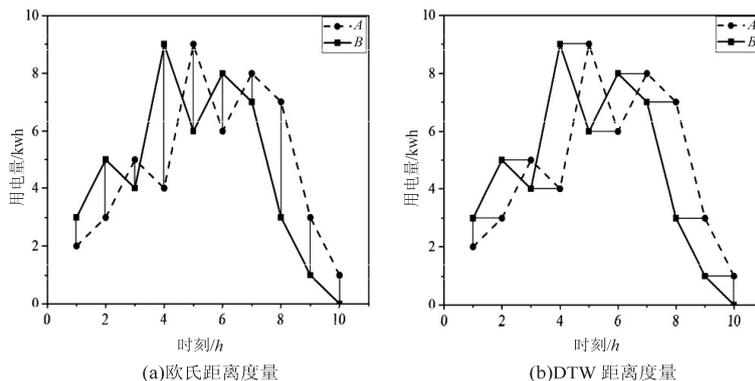


图 2 欧氏距离和 DTW 距离度量对比

1.3 使用 LB_Keogh 距离优化 DTW 距离

DTW-Kmedoids 算法使用 DTW 距离代替欧氏距离,可以更有效地度量时间序列的相似性,但是 DTW 的时间复杂度更高,计算代价较为昂贵,数据维数较大,逐个用 DTW 算法比较每一条曲线是否匹配非常耗时。因此引入 LB_Keogh 距离处理掉大部分不可能是最优匹配序列的序列,对于剩下的序列再使用 DTW 距离进行距离的度量。

假设有一条时间序列 $Q_{24} = \{q_1, q_2, \dots, q_{24}\}$ 。LB_Keogh 距离的定义包含两部分,第一部分定义 Q_{24} 的 $\{U_{24}, L_{24}\}$ 包络线,曲线 U_{24} 与曲线 Q_{24} 维数相同。根据公式 7 与公式 8 定义上下包络线,如图 3 所示。曲线 $U_{24} = \{u_1, u_2, \dots, u_{24}\}$ 为曲线 Q_{24} 的上包络线,曲线 $L_{24} = \{l_1, l_2, \dots, l_{24}\}$ 为曲线 Q_{24} 的下包络线。曲线 U_{24}, Q_{24}, L_{24} 在时刻上为一一对应关系索引值均为 i 。

$$u_i = \begin{cases} \max(q_{i-r}:q_{i+r}), & i-r > 0, i+r < 24 \\ q_i, & i+r \geq 24 \end{cases} \quad (7)$$

$$l_i = \begin{cases} \max(q_{i-r}:q_{i+r}), & i-r > 0, i+r < 24 \\ q_i, & i+r \geq 24 \end{cases} \quad (8)$$

式中, r 是上下包络线与中心曲线偏移的时间窗口,可以自行定义。图 3 中取 $r=2$ 对中心曲线进行包络线构建。

第二部分为:LB_Keogh 距离计算如公式 9 所示,图像描述如图 3 阴影部分所示。

$$LB_{Keogh} = \sqrt{\sum_{i=1}^{24} \begin{cases} (C_i - U_i)^2, & C_i > U_i \\ (C_i - L_i)^2, & C_i < L_i \\ 0, & C_i = L_i \end{cases}} \quad (9)$$

式中, LB_{Keogh} 表示曲线 C 与曲线 Q 的 LB_Keogh 距离。

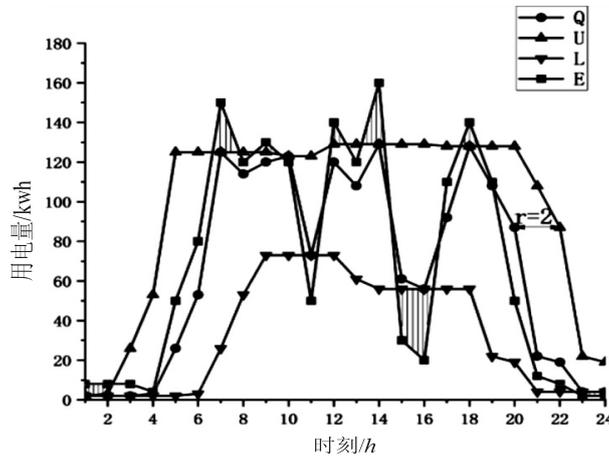


图3 上下包络线示意图

1.4 采用纵向滑动时间窗口降维

该文的数据来源为某高校电力数据采集系统采集的数据,考虑到高校建筑的用电习惯,每种类型的建筑物用电时间段不同,不必计算所有采样点的数据,因此引入纵向时间窗口,通过匹配特定时间段的采样点,丢弃其余采样点,最终达到降低计算量的目标。

设置时间窗口 $(tw, tw + ti)$, $ti \in (0, 24 - k)$ 。 tw 表示第 tw 个时刻, ti 表示时刻增量。例如,图 3 所示曲线 Q 用电主要集中在 4:00-22:00。所用到的有效采样点有 19 个,其余 5 个采样点所记录的数据对聚类没有太大影响,可以直接舍弃。

1.5 DLT-Kmedoids 算法步骤总结

DLT-Kmedoids 算法以准则函数最小为目标优化新的聚类中心,消除了算法对于孤立点的敏感性。准则函数如公式 10 所示:

$$SSE = \sum_{i=1}^k \sum_{Z \in C_{c_i}} D(Z, O_{c_i}) \quad (10)$$

式中, SSE 为误差平方和,在此作为准则函数值, C_{c_i} 为第 c_i 个簇, O_{c_i} 为第 c_i 个簇的聚类中心曲线, Z 为当前

簇中除去中心曲线的样本曲线, $D(\cdot)$ 为两条样本曲线的 DTW 距离, K 为最终聚类数。

DLT-Kmedoids 算法聚类中心曲线选取步骤如下:

Step1: 在总体 N 条样本曲线中选取任意 R 条曲线作为聚类中心曲线。

Step2: 使用时间窗口降维。

Step3: 在时间窗口内计算其余样本曲线与 K 条中心曲线的 LB_Keogh 距离,在 LB_Keogh 距离内计算 DTW 距离,按照与中心曲线距离最近原则,将其余 $N - R$ 条样本曲线分配到 K 条聚类中心曲线代表的簇中,从而实现初始聚类。

Step4: 针对第 c_v 个类中除中心曲线的其他样本曲线,计算当其作为新的中心曲线时的准则函数值,遍历所有曲线,选取准则函数值最小时的样本曲线为新的聚类中心曲线。

Step5: 重复 Step2 和 Step3,直到所有簇中的中心曲线不再发生变化。

Step6: 最终确定 K 个聚类中心。

2 评价指标

该文采用轮廓系数 (Silhouette Coefficient) 衡量聚类效果好坏,计算公式如公式 11 和公式 12 所示。

$$S(ki) = \frac{db(ki) - da(ki)}{\max\{da(ki), db(ki)\}} \quad (11)$$

$$SC = \frac{\sum_{i=1}^K S(ki)}{K} \quad (12)$$

式中,da 表示某个样本与其所在簇内其他样本的平均距离,体现簇内凝聚度。db 表示某个样本与其他簇样本的平均距离,体现簇间分离度。S(ki) 为第 ki 个簇的轮廓系数,SC 为所有簇轮廓系数的平均值,用来衡量整体轮廓系数。

3 实验

该文以建筑物实际日负荷用电曲线为基础数据,分别采用 Kmedoids 算法^[19]、DTW-Kmedoids 算法、DLT-Kmedoids 算法以及 FCM 算法^[20]对不同建筑物真实用电负荷曲线进行挖掘。

3.1 实验数据

以某高校五类建筑物(建筑物类型包含图书馆、行政楼、体育馆、教学楼以及食堂)共 11 栋建筑物电力数据采集系统记录的日负荷曲线为研究对象,数据粒度为 1h/点,每天共有 24 个采样点,经过数据预处理后,共得到 2 080 条有效样本曲线。

3.2 手肘法确定聚类数

使用手肘法确定聚类数^[21],选取聚类数 2~10 进行实验,使用该文提出的算法时设置衡量 LB_Keogh 距离的时间步长 r=2,滑动时间窗口设置为 8:00-20:00。结合图 4 和图 5 看,针对本数据集,四种算法在聚类数 K=2 时轮廓系数 SC 最大,但是误差平方和 SSE 也处于较高位置,可以判定当 K=2 时四种算法均陷入了局部最优值点。综合考虑误差平方和和轮廓系数,该文选择四种算法聚类数 K=5 进行后续实验。

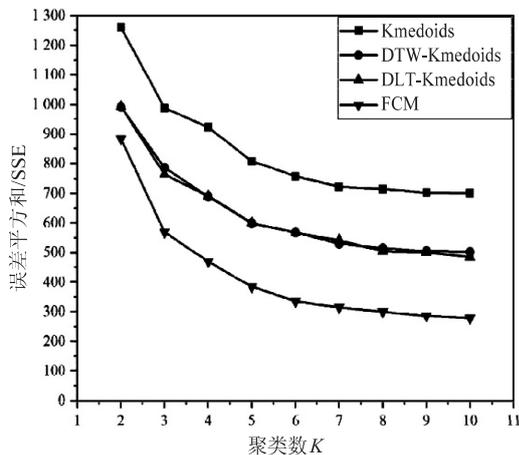


图 4 误差平方和对比

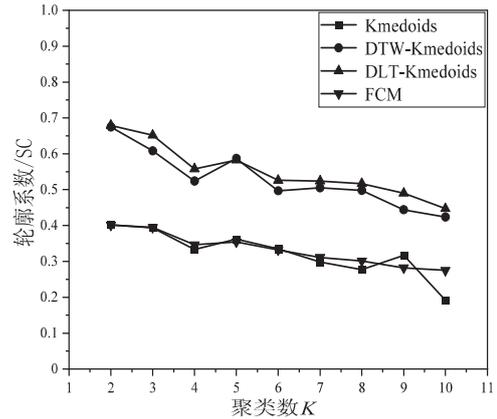
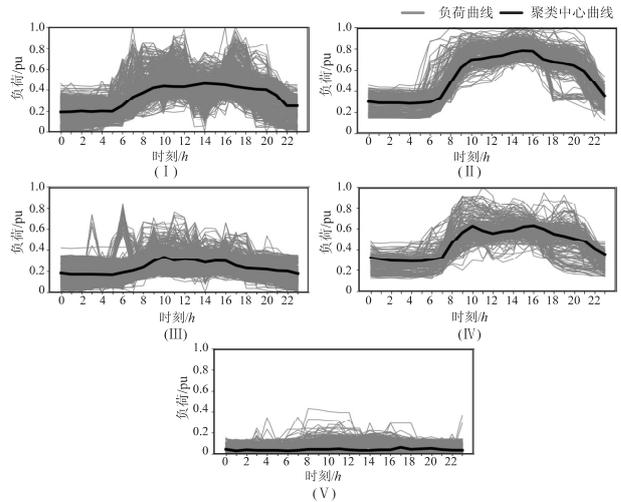


图 5 轮廓系数对比

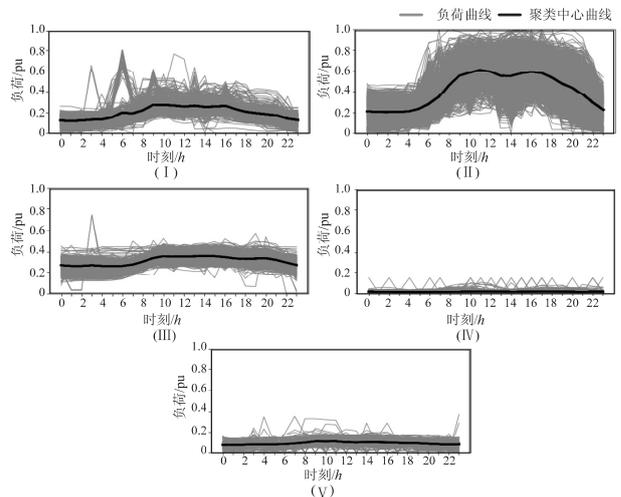
3.3 聚类效果分析

通过手肘法选出最佳聚类 K=5,设置以上四种算法迭代次数为 40 进行实验,结果如图 6 所示。

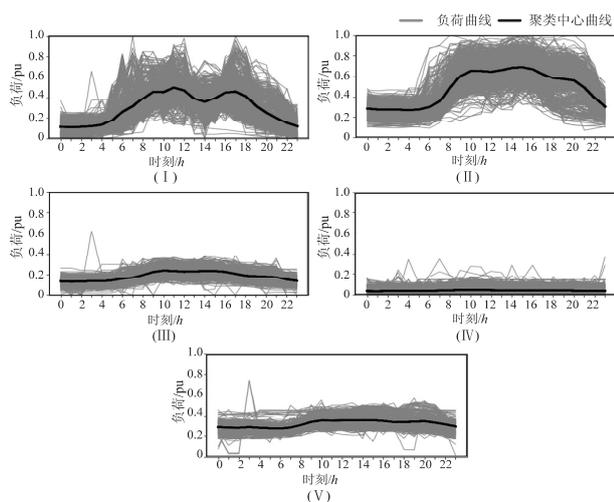
算法把数据集分为五类,第(I)类曲线为双峰型曲线,对比真实数据看,其中包含行政楼和食堂两种类型建筑,行政楼峰值出现在 9:00 和 14:30,食堂峰值出现在 11:00 和 17:00,由于算法引入了 DTW 距离消除了横向时间的差异,所以将这两类建筑的双峰曲线归



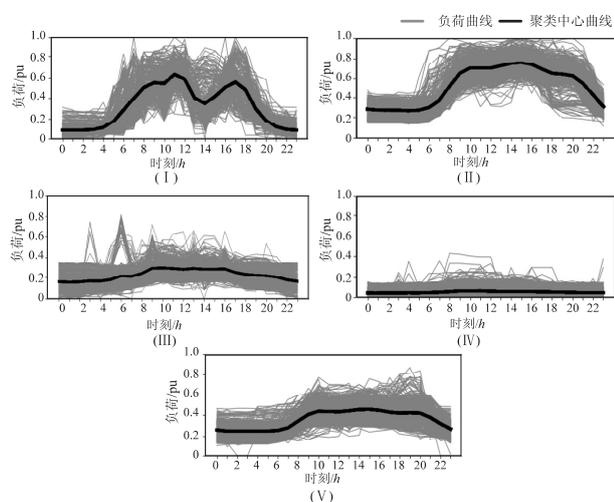
(a) Kmedoids 算法结果



(b) DTW-Kmedoids 算法结果



(c) DLT-Kmedoids 算法结果



(d) FCM 算法结果

图 6 不同聚类算法聚类结果对比

为一类。第(II)类用电曲线从7:00开始到20:00均维持在一个较高水平,在21:00用电量开始下滑。这一类为图书馆和教学楼这种用电比较单一的建筑,主要用电来源于空调和照明。学生在11:00与18:00去食堂吃饭或者休息造成用电量下滑,休息完后一部分人回来继续自习,所以在14:30和19:30会出现小峰值。在21:00过后大部分人选择回宿舍休息,因此教学楼与图书馆用电开始断崖式下滑。第(III)类用电在8:00开始达峰全天维持在同一水平,该类为体育馆用电,体育馆用电较为粗放用电波动不大。第(IV)类用电基本为0是假期建筑物用电的体现。第(V)类为假期教学楼和图书馆用电曲线,在假期中教学楼与图书馆仍然会有小部分学生按照平常习惯性自习,因此总体来说第(V)类用电形态与第(II)类相似,但是用电量级明显减少。

针对第(I)类建筑中的食堂,高校可以通过合理错峰排课,引导学生错峰用餐。对于行政楼来说,可以调整各部门工作时间表,合理安排值班人员。针对第(II)类与第(V)类建筑,管理者应当将学生尽可能安

排在同一区域,其他区域关闭电,以达到科学用电的目的。针对第(III)类建筑,体育馆属于专用场所,在某一区域无人使用时,应当关闭用电以节约用电量。针对第(IV)类节假日用电情况,管理者应当关闭用电以节约设备待机用电量。

相比于传统算法,使用DTW距离度量能够把总用电量相同的双峰形曲线和单峰曲线准确聚为第(I)类和第(II)类。相比于DTW-Kmedoids算法,DLT-Kmedoids算法较为清晰地展现出第(V)类与第(II)类有相似的峰谷特征。相比于FCM算法,模糊聚类将更多的双峰曲线错误地划分进了第(III)和第(V)类中,导致类别之间的界限并不明显。从评价指标看,文中方法相比于FCM算法轮廓系数更高,划分更加紧密。

3.4 聚类算法时间

设置迭代次数为40, LB_Keogh距离的时间步长 $r=2$,滑动时间窗口设置为8:00-20:00。对文中算法和DTW-Kmedoid算法进行比较。算法时间对比如图7所示。

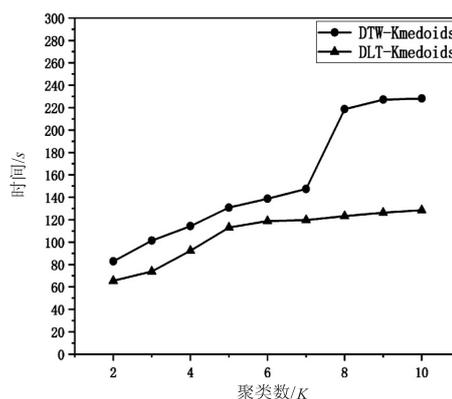


图 7 算法时间对比

通过图4、图5、图7, DLT-Kmedoids算法与DTW-Kmedoids算法对比看SSE基本不变, SC略有上升并且平均降低了27%的算法迭代时间,提升了算法的快速性。

4 结束语

针对Kmedoids算法对于时间序列曲线聚类精度低, DTW-Kmedoids算法运算复杂度高的问题,提出了一种基于双距离度量与时间窗口融合的时间曲线聚类算法。算例结果表明:

(1)在实际用电负荷数据集中, DLT-Kmedoids算法能够较为准确地划分聚类类别,相比于Kmedoids算法,有效提升了聚类准确度。

(2)相较于DTW-Kmedoids算法, DLT-Kmedoids算法在基本不影响准确度的情况下降低了27%的算法迭代时间,提升了算法的快速性。

(3)在高校用电管理方面,体育馆和图书馆这类大型公共建筑采用粗放型管理模式,具有较大的节能空间。

(4)高校食堂和行政楼用电负荷曲线具有典型双峰形态,应当引导学生错峰上课,行政工作人员错峰就餐,以达到去峰填谷的作用。

(5)假期高校图书馆,教学楼仍然有人分散自习,可以划分单独自习区域以节约照明和空调用电量。

综上所述,DLT-Kmedoids 算法能够较好地辅助高校管理者分析电力用户的用电特性,从而对用电用户更为精细化划分,制定合理的用电策略,以达到节约用电的目的。该方法目前主观地根据用电负荷典型曲线制定用电策略。下一步将根据用电负荷典型曲线构建异常检测模型,进行异常数据检测,针对高校建筑进一步优化节能策略。

参考文献:

- [1] QUILUMBA F L, LEE W J, HUANG H, et al. Using S-mart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities [J]. IEEE Transactions on Smart Grid, 2015, 6(2): 911-918.
- [2] 王毅, 张宁, 康重庆, 等. 电力用户行为模型: 基本概念与研究框架 [J]. 电工技术学报, 2019, 34(10): 2056-2068.
- [3] 李海林, 张丽萍. 时间序列数据挖掘中的聚类研究综述 [J]. 电子科技大学学报, 2022, 51(3): 416-424.
- [4] 孟建良, 刘德超. 一种基于 Spark 和聚类分析的辨识电力系统不良数据新方法 [J]. 电力系统保护与控制, 2016, 44(3): 85-91.
- [5] 高小力, 张智博, 田启明, 等. 基于 HS-Clustering 的风电场机分组功率预测 [J]. 现代电力, 2017, 34(3): 12-18.
- [6] 黄薇, 温蜜, 张照贝. 考虑用户用电行为聚类的电力负荷预测方法 [J]. 计算机仿真, 2022, 39(12): 148-153.
- [7] 曾四鸣, 李铁成, 李顺, 等. 基于改进型密度峰值算法的电力负荷聚类分析 [J]. 科学技术与工程, 2022, 22(25): 11032-11040.
- [8] LIAO T W. Clustering of time series data—a survey [J]. Pattern Recognition, 2005, 38(11): 1857-1874.
- [9] 党倩, 崔阿军, 尚闻博, 等. 采用欧式形态距离的负荷曲线近邻传播聚类方法 [J]. 西安交通大学学报, 2022, 56(1): 165-176.
- [10] 孙园, 李秋雨, 黄冬梅, 等. 基于粒计算和双尺度相似性的负荷曲线集成聚类算法 [J]. 电力建设, 2022, 43(5): 117-126.
- [11] 李克明, 江亚群, 黄世付, 等. 基于 DTW 距离和聚类分析的配电台区低压拓扑结构辨识方法 [J]. 电力系统保护与控制, 2021, 49(14): 29-36.
- [12] 白雅玲, 周亚同, 刘君. 基于深度卷积嵌入聚类的日负荷曲线聚类分析 [J]. 电网技术, 2022, 46(6): 2104-2113.
- [13] 邓春宇, 吴克河, 张玉天. 基于近邻相似度图聚类的用户用电行为分 [J]. 计算机工程与设计, 2021, 42(7): 1933-1939.
- [14] 陈乾, 胡谷雨. 一种新的 DTW 最佳弯曲窗口学习方法 [J]. 计算机科学, 2012, 39(8): 191-195.
- [15] CAO D, LIU Y, LI G, et al. An algorithm on time series matching based on LB_Keogh and early abandon [J]. International Journal of Advancements in Computing Technology, 2012, 4(15): 280-286.
- [16] 庞新生. 缺失数据插补处理方法的比较研究 [J]. 统计与决策, 2012(24): 18-22.
- [17] 王帅, 杜欣慧, 姚宏民, 等. 面向含多种用户类型的负荷曲线聚类研究 [J]. 电网技术, 2018, 42(10): 3401-3412.
- [18] LEMIRE D. Faster retrieval with a two-pass dynamic-time-warping lower bound [J]. Pattern Recognition, 2008, 42(9): 2169-2180.
- [19] 曾建邦, 张月娅, 张壮, 等. 基于动态 k 值 K-means++ 聚类的电动汽车动力电池电压不一致故障识别方法 [J]. 中国科学: 技术科学, 2023, 53(1): 28-40.
- [20] 柳劲松, 王颖, 熊春晖, 等. 基于日负荷曲线的用电模式提取与聚类分析 [J]. 电工电能新技术, 2023, 42(5): 57-63.
- [21] 陈聿, 田博今, 彭云竹, 等. 联合手肘法和期望最大化的高斯混合聚类电力系统客户分群算法 [J]. 计算机应用, 2020, 40(11): 3217-3223.